Journal of Epidemiology and Biostatistics (2001) Vol. 6, No. 5, 393–407

Repeatability of sexual history in longitudinal studies on HPV infection and cervical neoplasia: determinants of reporting error at follow-up interviews

NF SCHLECHT[1], EL FRANCO[1], TE ROHAN[2], SK KJÆR[3], MH SCHIFFMAN[4], A-B MOSCICKI[5], AND SW DUFFY[6], FOR THE MARSH INTERNATIONAL COLLABORATIVE GROUP[7]

[1] *Departments of Epidemiology and Oncology, McGill University, Montreal, Canada*

[2] *Department of Epidemiology and Social Medicine, Albert Einstein College of Medicine, NY, USA*

[3] *Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark*

[4] *Environmental Epidemiology Branch, National Cancer Institute, Rockville, USA*

[5] *Department of Pediatrics, University of California at San Francisco, USA*

[6] *Imperial Cancer Research Fund, London, UK*

[7] *Investigators in the Meta-Analysis of Repeatability of Sexual History (MARSH) International Collaborative Group (in order of size of study): Copenhagen Cohort (Denmark): SK Kjær, EI Svare; National Institutes of Health (NIH) Guanacaste Project (Costa Rica): M Schiffman, A Hildesheim, R Herrero, MC Bratti, AC Rodriguez; Ludwig-McGill Cohort (São Paulo): EL Franco, LL Villa, A Ferenczy, TE Rohan; University of Toronto Student Cohort: TE Rohan, DG Harnish, EL Franco; McGill/Concordia University Student Cohort (Montreal): EL Franco, H Richardson, F Coutlée, A Ferenczy, P Tellier; San Francisco State University (SFSU) Cohort of Adolescents and Young Women: AB Moscicki, J Palefsky, S Shiboski, TM Darragh, N Benowitz; Statistical analysis: NF Schlecht, EL Franco, TE Rohan, SW Duffy.*

Background Misclassification of sexual history due to faulty recall or reporting bias may be the reason for variability in the association between sexual history and human papillomavirus (HPV) infection seen in studies conducted in different geographical areas. This study aimed to assess the repeatability of questionnaire information on sexual-history variables and their correlates, using information from repeat interviews by six international prospective cohort studies.

Methods The pooled dataset included over 14 775 women interviewed on two separate occasions, of whom 5690 returned for a third interview. At each return visit women were re-asked questions on age at first intercourse and number of sexual partners. The six cohorts originated from studies in Denmark, Costa Rica, San Francisco, Toronto, Montreal and São Paulo.

Results Exact agreement between age at first intercourse recalled on separate occasions ranged from 60–85%, whereas exact recall rates for number of sexual partners were substantially lower and more study-dependent, varying between 20% and 77%. The intraclass correlation coefficients gauging the degree of repeatability in responses ranged from 0.68 to 0.97 for age at first intercourse and 0.08 to 0.94 for number of sexual partners. Age, ethnicity, education and cohort membership were the strongest predictors of reporting error for both sexual history markers, although study design characteristics also seemed to play a role. HPV infection status seemed to influence recall of number of partners, but not age at first intercourse.

Conclusions Information on sexual behaviours is not reliably collected in epidemiological studies of sexually transmitted diseases, which may influence the magnitude of relative risk estimates.

Keywords misclassification, sexual behaviour, human papillomavirus, cervical neoplasia, cohort studies, information bias.

Introduction

Epidemiological evidence suggests that sexual history plays a pivotal role in the transmission of genital human papillomavirus (HPV) infection, the latter being a central event in the development of cervical cancer[1]. An independent association between sexual activity and cervical-cancer risk has been the hallmark of most epidemiologic studies conducted in the past 30 years[2], but this relationship disappears upon adjustment for HPV infection status, reflecting the intermediate position of HPV infection in the causal pathway between sexual activity and cervical neoplasia[3–5]. However, measurement error still

Correspondence to: EL Franco, Department of Oncology, McGill University, 546 Pine Avenue West, Montreal, QC, Canada H2W 1S6.

Received 28 February 2001
Revised 18 June 2001
Accepted 5 July 2001

constrains our ability to evaluate key steps in the natural history of cervical cancer. In particular, a clear understanding of the nature and strength of the relationship between specific markers of sexual behaviour and risk of HPV infection is still lacking, as results of studies of the association between sexual history and HPV infection have yielded conflicting results[6–14].

While this issue may have been due to some extent to misclassification of HPV infection, it is also plausible that misclassification of sexual activity information due to faulty recall or reporting bias may contribute to attenuate the strength of the association between sexual history and HPV infection[15]. Few studies have investigated the effect of misclassification due to reporting error in sexual behaviour[16–18], or the determinants of such errors and the effect of disease status on the quality of the information[19]. Difficulty in assessing the degree of misclassification has been the result primarily of small study sizes and the absence of sentinel questions asked at repeated interviews[20,21].

In this international collaborative study we combined data from six cohort studies investigating the natural history of HPV infection and cervical neoplasia[5,14,19,22–24] to assess the extent of error in the measurement of sexual activity markers and the characteristics that contribute to such misclassification. Questions on the number of sexual partners and age at first sexual intercourse were collected at return visits by each of the studies, to evaluate the repeatability of sexual history. HPV infection status by viral DNA testing and Pap cytology testing for cervical lesions were obtained at the baseline visit to determine outcome status upon entry.

Methods

Cohort studies

The six cohorts originated from both population- and clinic-based studies in Denmark (Copenhagen Cohort), Costa Rica (National Institutes of Health Guanacaste Project), San Francisco (San Francisco State University (SFSU) Cohort of Adolescents and Young Women), Toronto (University of Toronto Student Cohort), Montreal (McGill-Concordia University Student Cohort) and São Paulo (Ludwig-McGill Cohort). Table 1 describes design characteristics for these prospective cohorts in order of study size. Recruitment of subjects for each study was done either through student health clinics at McGill and Concordia Universities[14], the University of Toronto[19], and SFSU[22]; through community clinics in San Francisco[22] and São Paulo[24]; or by random selection based on a population census in Costa Rica[23] and Denmark[13,25]. All participants in each study signed informed consent forms. Study protocols and informed consent documents were approved by the institutional review boards of all institutions with which the authors are affiliated.

Women eligible to participate in the studies were not pregnant at enrolment (except in Denmark, where a random sample of all women in the population was recruited); had an intact uterus, and were not receiving treatment for cervical disease at enrolment. Women in the São Paulo study were scheduled for return visits every 4 months for the first year and twice yearly thereafter. Women recruited through the student health clinics at the University of Toronto, and the McGill and Concordia Universities in Montreal, were scheduled for return visits every year and every 6 months respectively. Although interval periods between return visits for the SFSU cohort were in fact 4 months in duration, the sentinel questions on sexual history were not repeated until several visits into the study, an average of 5 years later (Table 1).

Baseline and referral procedures for return visits also varied among cohort studies. In the SFSU cohort, women were tested for HPV in an initial screening visit before enrolment. Women found to be HPV positive, plus a random sample of those with negative HPV results, were then invited to an interview where questions on sexual history were posed[26]. In the Guanacaste cohort, women with evidence of cervical lesions at the baseline visit were first referred for colposcopy 6–10 weeks later, along with a random sample of the remainder selected at baseline. Women showing evidence of low-grade lesions were subsequently followed every 6 months. Those with atypical cytology results, testing positive for HPV, with five or more sexual partners and women with normal results were screened every year thereafter[23]. In the interest of comparability, the second of the two 6-month follow-up visits for women with low-grade lesions was used (referred to as the first follow-up interview hereafter); responses given at the colposcopic exam interview were considered separately, but not included in the analyses.

Study variables

Information on sexual history and demographic characteristics was obtained through interviews, conducted in the location's primary language, for all studies except the McGill-Concordia University Student Cohort, which used self-administered questionnaires. The sentinel questions posed at each return visit were asked in a comparable manner by each study: 'How old were you when you first had sexual intercourse?' and 'With how many partners have you had sexual intercourse over your lifetime?' Furthermore, at return visits, women were asked how many new sexual partners they had had since the previous visit. Responses to these questions were used to gauge the level of reporting error for sexual history.

Reporting error for age at first intercourse was calculated by subtracting the original response from the one obtained at the repeat interview. Subjects who had sexual

Table 1 Summary of study design characteristics for each participating cohort

Study designation and location	Total no. of interviews	Start of data collection	Time to 2nd interview	Time to 3rd interview	No. with two interviews	No. with three interviews	Sample selection	Interview procedure	HPV testing method	Cervical lesion classification[a]
Copenhagen Cohort, Denmark	2	1991	2 years		8656		Population based	Face-to-face	PCR	TBS
NIH Guanacaste Project, Costa Rica	3[b]	1993	6 to 10 weeks	6 months to 1 year	2139	2887	Population based	Face-to-face	Hybrid capture[c]	WHO/TBS
Ludwig-McGill Cohort, São Paulo, Brazil	8	1993	4 months	4 months	2188	1957	Multiple clinics	Face-to-face	PCR	TBS
University of Toronto Student Cohort, Toronto,Canada	3	1991	1 year	1 year	1036	483	Single clinic	Face-to-face	PCR	WHO
McGill-Concordia University Student Cohort, Montreal, Canada	5	1996	6 months	6 months	548	363	Multiple clinics	Self-administered	PCR	TBS
SFSU Cohort of Adolescents and Young Women, San Francisco, USA	2	1989	5 years[d]		208		Multiple clinics	Face-to-face	ViraPap/ Profile Test/PCR	WHO

[a] WHO = WHO classification based on the cervical intra-epithelial neoplasia (CIN) designation, TBS = the Bethesda System of cytological diagnoses based on atypical squamous cells of undetermined significance (ASCUS) and squamous intra-epithelial lesion (SIL) designations.
[b] Third interview for Guanacaste represents the first follow up interview following colposcopic exam, where the 1-year interval interview was used when two 6-month follow up visits were available.
[c] Retesting using PCR based methods has been completed for the Guanacaste cohort, although results were not available at the time of analysis for this study.
[d] Second interview for SFSU represents the first follow up interview where sentinel questions on sexual history were asked again, although this does not correspond to the first return visit in that study.

intercourse for the first time between the baseline interview and a return visit were excluded from the analysis. Responses at the return visits were used to derive the lifetime number of sexual partners at the baseline visit by subtracting the reported number of new partners since previous visit(s) from the total number reported at the same follow-up visit. These values were compared with the original measure in the same fashion as age at first intercourse.

Common questions on demographic characteristics for all cohorts were also used: ethnicity, level of schooling, marital status, smoking history and reporting of sexually transmitted diseases (STD) prior to enrolment. History of STDs was measured by such questions as: 'Have you ever had an STD?' and 'Has your physician ever told you that you had an STD?' followed by specific questioning on types of STD.

Cervical cytology and HPV testing

Conventional Pap smears were used in all cohort studies to obtain a presumptive diagnosis of cervical lesions. The system of cervical cytology classification varied by cohort (Table 1). For studies using the Bethesda classification, cervical lesion status was defined as positive for those individuals with a diagnosis of at least low-grade squamous intraepithelial lesion (LSIL), whereas for studies using the World Health Organisation (WHO) scheme cervical lesion status was defined as positive for women with at least cervical intraepithelial neoplasia Grade 1 (CIN-1 or equivalently, mild dysplasia or dyskariosis).

HPV infection status was determined at each visit (all cohorts except Costa Rica, where results at enrolment only were available) by testing for viral DNA using either polymerase chain reaction (PCR) protocols based on consensus, or general primer amplification of conserved viral genome regions, HPV Profile Test (starting the second year, San Francisco) or the Hybrid Capture assay (Costa Rica). The ViraPap assay was initially used in the San Francisco study to select women to be enrolled into the cohort.

Statistical analysis

Mean differences in responses and intraclass correlation coefficients (ICC) were calculated using values for the two sexual behaviour sentinel questions reported at multiple repeat visits, to evaluate the agreement between the original and repeat responses and the degree of reporting error. ICCs for the combined analyses were pooled by weighted average, with weights proportional to the inverses of the variances[27]. Before pooling, ICC's were first normalised using the Fisher z-transformation. The transformation is designed for the product–moment correlation coefficient and the transform of the latter will have a standard error of unity. Since the standard error (SE) of the ICC is different from that of the product–moment coefficient, this could not be assumed; therefore the SE of the transformed ICCs were estimated by a Taylor series approximation. The pooling method provides estimates of the overall ICC, SE and a significance test for heterogeneity among studies. ICC values close to 1.0 indicate that a larger proportion of the variance observed is due to variability between rather than within subjects[28].

In addition, comparison of the same baseline measures for age at first intercourse and lifetime number of partners was also done, using a categorical form of the responses based on typical cutpoints used in epidemiologic studies of cervical cancer. For each cohort, the degree of discrepancy across categories of these two variables between the first two interviews was estimated by the κ statistic[28].

Logistic regression with a random effects approach was employed to determine which demographic factors were associated with reporting error for both sexual activity markers. Two binary forms of each variable were used to gauge reporting error:

- Any deviation from the original response
- Extreme variation, defined as responses differing by more than 2 years for age at first intercourse, or by more than two partners for lifetime number of sexual partners.

Random effects modelling was carried out on the assumption of similar trends in the effect of risk factors on reporting error across each of the cohorts, allowing for unexplained variability among cohorts. Estimates of effect were presented as odds ratios (OR) with 95% confidence intervals (CI).

Results

For all cohort studies, a total of 14 775 women were interviewed both at baseline and at the first return visit. Of these women, 5690 and 2045 had third and fourth interviews, respectively. The number of subjects with responses at a second visit represented 78% of the initial numbers for Denmark, 29% for Costa Rica (including the first recorded follow-up visit after colposcopy), 86% for São Paulo, 68% for Toronto, 86% for Montreal and 23% for San Francisco.

Table 2 describes the distribution of selected demographic and sexual history characteristics for each of the cohorts. The ethnic admixture in the Copenhagen cohort was minimal, consisting almost entirely of Caucasians. However, due to the level of admixture in the Costa Rica population, which includes women of African or Asian origin as well as of European or Native American stock, classification of ethnicity was not possible. The 'other'

Table 2 Descriptive statistics of socio-demographic characteristics, sexual history, HPV status and cervical lesions taken at baseline and follow-up interviews for all cohorts

Variable[a]	Denmark	Costa Rica	São Paulo	Toronto	Montreal	San Francisco
Level of schooling (%)						
Primary	6	56	81	-	-	-
Secondary	23	23	16	-	-	55
College	71	14	3	100	100	44
Ethnicity (%)						
Caucasian	100	-	64	90	79	72
African origin	-	-	13	1	3	16
Asian origin	-	-	1	8	10	4
Other	-	-	22	1	8	8
Marital status (%)						
Single	47	25	10	75	77	89
Married	52	67	82	19	17	3
Separated/divorced	1	8	8	6	2	-
Smoking history (%)						
Never	46	90	27	80	64	33
Ex-smoker	10	5	37	11	12	37
Current smoker	44	5	36	10	23	30
% Positive STD history	34	4	9	17	19	41
% HPV positive at baseline	15[c]	18	14	13	17	46
% SIL at baseline	2	8	2	1	2	6
Mean[b] age (years) at baseline interview	24.9 (23-27)	36.9 (25-46)	32.9 (26-39)	23.2 (21-25)	22.5 (20-24)	18.0 (17-19)
Mean[b] age (years) at first intercourse	16.2 (15-18)	15.6 (15-19)	17.9 (15-20)	18.3 (17-20)	17.2 (16-19)	15.0 (14-16)
Mean[b] no. of sexual partners at baseline	9.8[d] (4-12)	2.3 (1-3)	2.8 (1-3)	4.1 (1-5)	6.2 (2-8)	6.1 (2-8)
Mean[b] no. of sexual partners at first follow-up	11.5[d] (5-17)	2.3[e] (1-3)	2.9 (1-3)	5.8 (2-6)	6.7 (2-8)	11.1 (5-13)

[a] Percentages (%) correspond to values collected at the baseline interview among women who returned for a follow-up visit. Unknown categories included in the calculation of percentages are not presented.
[b] Mean and interquartile range markers (25–75%).
[c] Prevalence of HPV based on a random sample of women in the Denmark population cohort.
[d] Calculation for no. of sexual partners based on midpoint values from categories in the Denmark cohort.
[e] Response taken at the first interview following colposcopic examination.

category in the São Paulo cohort includes mostly women classified as mulatto or mestizo, a mix of both European and Afro–Brazilian ethnicities.

Due to recruitment and referral procedures in the Guanacaste and SFSU cohorts, the positivity for HPV infections and squamous intra-epithelial lesion (SIL) are higher than expected for the respective population areas. The mean age at first intercourse did not vary much among cohorts, although the mean number of sexual partners at the baseline visit was higher in the North American university-based cohorts than in the South and Central American cohorts. Calculation of the number of sexual partners in the Copenhagen cohort was based on midpoint values for categorical measures of five or more partners (5–9, 10–14, 15–19, 20–24, 25–29, 30–34, 35–39 and 40+).

Table 3 illustrates the agreement between responses from pairs of interviews for age at first intercourse. The within-subject variability was largest for women in the SFSU cohort (exact agreement = 63%). However, judging by the frequencies of discrepant responses and by the mean differences for all subjects (i.e. those with and without discrepant responses), repeated responses were more discrepant in magnitude for women in the South and Central American cohorts. On the other hand, the Montreal cohort seemed to have on average the most numerically disparate repeat responses (1.91 years ± 0.31) when the analysis was restricted to those with discrepant values. Recall was similar for the referral and first follow-up visits in the Guanacaste cohort, although the range of responses seemed greater at the former — as indicated by a lower degree of correlation.

There was considerably more variation across cohorts with respect to the repeatability of the information for lifetime number of sexual partners (Table 4). The disagreement increased with the interval between interviews, as did the average difference for those giving discrepant responses. Estimates of differences in lifetime number of sexual partners were not derived for Denmark because responses were not collected as continuous variables at the relevant interviews.

Table 5 shows the variability in responses for the above two indicators of sexual activity, pooled for all cohorts combined and stratified by HPV and lesion status. Subjects with unknown HPV or lesion status were included in the overall combined analysis. Estimates for number of sexual partners from Denmark were again excluded. Responses provided at the first follow-up interview after colposcopic exam were used for the Guanacaste cohort. The exact agreement for number of sexual partners was somewhat worse for HPV-positive than for HPV-negative women. The degree of heterogeneity among cohorts was significant across all levels of stratification except among SIL positive subjects. Removing observations with large discrepancies in reported number of partners and age at first intercourse had a negligible effect on ICC estimates.

When compared in categorical form (Table 6), which is typically the way measures of sexual activity are analysed in epidemiologic studies of cervical cancer, the degree of variability in age at first intercourse and number of sexual partners between baseline and first follow-up visits decreased considerably for all cohorts. For age at first intercourse, the highest between-interview agreement was seen for the two university cohorts, as indicated by the > 90% rates for perfect agreement (identical coding in both interviews) and by the high κ statistics. The ranking of cohorts for the degree of variability in categorical form was approximately the same as that resulting with original coding, for both age at first intercourse and lifetime number of sexual partners. We also estimated disagreement using only the total number of sexual partners reported at each visit. Instances of women reporting fewer total partners at return interviews than at baseline represented 8% of the Denmark cohort, 13% for Guanacaste, 11% for São Paulo, 6% for Toronto, 10% for Montreal and 12% for San Francisco.

Two series of logistic regression models for variation in both age at first intercourse and lifetime number of sexual partners are shown in Table 7. The probability of reporting error increased with age, lower education and ethnicity of African or other ancestry (non-Caucasian, non-Asian), with the magnitude of the associations being greater when age at first intercourse was defined for more extreme differences in reporting (> 2 years discrepancy). Cohort membership was also an important predictor of accurately reporting age at first intercourse, with lowest discrepancy rates for the Montreal cohort, for any error, and for the Denmark cohort, for extreme variation. Even without adjustment for cohort membership there was a considerable degree of explanatory value contributed by other variables included in the model (data not shown). Interval time seemed to affect any reporting error, but was not a predictor of extreme variation in responses. HPV or lesion status was not associated with reporting error, regardless of the degree of variation. The main independent determinants of any reporting errors for number of sexual partners were age, education, cohort membership, HPV status, baseline number of sexual partners and new sexual partners. Results for extreme variation in responses for lifetime number of partners were comparable, except for the contribution of STD history, marital status and a change in the direction of the association with HPV status: positive for any variation and negative for extreme variation.

Discussion

The degree of precise recall for age at first intercourse at repeat interviews ranged between 60% and 85% for the collaborating cohorts, whereas exact recall rates for number of sexual partners were substantially lower and more study-dependent, varying between 20% and 77%. Most of the variance observed was attributable to variations between subjects in the cohorts, rather than between visits for the same subjects. The difference in magnitude of the ICCs between collaborating cohorts did not seem to correlate with differences in interval time over which subjects had to recall information. This may be a reflection of the study populations, or designs and interview procedures, rather than recall time. Kunin and Ames[20] observed similar recall rates when comparing logbook entries made daily and monthly by women attending family clinics. Inter-subject variance only increased with interval period for number of sexual

Table 3 Summary statistics for the degree of repeatability in age at first intercourse between interviews for all cohorts

Interview comparison	Number of individuals (%) according to level of discrepancy between repeat and original measures[a]					Mean difference (SE)	
	≤-2	-1	0	1	≥2	All subjects[b]	Discrepant only[c]
Denmark							
1 and 2	220 (2.6)	1274 (14.9)	5646 (65.8)	1200 (14.0)	238 (2.8)	0.40 (0.01)	1.18 (0.01)
Costa Rica							
1 and 2	202 (5.8)	364 (10.4)	2129 (60.7)	487 (13.9)	200 (5.7)	0.68 (0.06)	1.80 (0.15)
1 and 3[d]	172 (6.2)	310 (11.2)	1702 (61.6)	411 (14.9)	167 (6.0)	0.67 (0.03)	1.73 (0.05)
São Paulo							
1 and 2	157 (7.2)	286 (13.1)	1342 (61.4)	263 (12.0)	138 (6.3)	0.65 (0.02)	1.68 (0.04)
1 and 3	131 (6.7)	247 (12.6)	1207 (61.7)	241 (12.3)	130 (6.6)	0.64 (0.03)	1.67 (0.05)
1 and 4	117 (6.4)	239 (13.1)	1091 (59.8)	243 (13.3)	133 (7.3)	0.67 (0.03)	1.68 (0.05)
Toronto							
1 and 2	27 (2.6)	156 (15.2)	732 (71.1)	106 (10.3)	8 (0.8)	0.36 (0.02)	1.24 (0.05)
1 and 3	13 (2.7)	69 (14.5)	346 (72.5)	41 (8.6)	8 (1.7)	0.34 (0.03)	1.23 (0.05)
Montreal							
1 and 2	8 (1.7)	37 (7.8)	400 (84.0)	23 (4.8)	8 (1.7)	0.30 (0.06)	1.91 (0.31)
1 and 3	4 (1.3)	22 (1.7)	265 (85.8)	11 (3.6)	7 (2.3)	0.26 (0.06)	1.80 (0.37)
1 and 4	2 (1.0)	9 (4.7)	165 (86.4)	10 (5.2)	5 (2.6)	0.20 (0.05)	1.50 (0.22)
San Francisco							
1 and 2	5 (2.4)	27 (13.0)	131 (63.3)	35 (16.9)	9 (4.3)	0.57 (0.09)	1.55 (0.21)

[a] Level of discrepancy calculated by subtracting original measure from repeat measure excluding subjects who reported having had no sexual intercourse at the baseline interview.

[b] Mean difference and SE based on absolute values using all subjects in each cohort with follow-up interviews.

[c] Mean difference and SE based on absolute values using only subjects with discrepant responses between baseline and follow-up interviews.

[d] Third visit corresponding to the first interview following colposcopic examination.

Table 4 Summary statistics for the degree of repeatability in lifetime number of sexual partners between interviews for all cohorts

	Number of individuals (%) according to level of discrepancy between repeat and original measures[a]					Mean difference (SE)	
Interview Comparison	**≤-2**	**-1**	**0**	**1**	**≥2**	**All subjects[b]**	**Discrepant only[c]**
Costa Rica							
1 and 2	146 (4.2)	265 (7.6)	2588 (73.8)	217 (6.2)	122 (3.5)	0.40 (0.07)	2.18 (0.36)
1 and 3[d]	151 (5.5)	255 (9.3)	2052 (75.0)	175 (6.4)	102 (3.7)	0.50 (0.03)	1.99 (0.08)
São Paulo							
1 and 2	77 (3.5)	179 (8.2)	1687 (77.3)	154 (7.1)	85 (3.9)	0.62 (0.11)	2.74 (0.47)
1 and 3	73 (3.7)	163 (8.3)	1473 (75.4)	165 (8.4)	81 (4.1)	0.70 (0.11)	2.83 (0.42)
1 and 4	69 (3.8)	173 (9.5)	1362 (74.9)	131 (7.2)	84 (4.6)	1.14 (0.36)	4.54 (1.41)
Toronto							
1 and 2	44 (4.5)	74 (7.6)	628 (64.3)	131 (13.4)	100 (10.2)	0.75 (0.05)	2.09 (0.12)
1 and 3	19 (4.5)	36 (8.5)	250 (58.7)	65 (15.3)	56 (13.1)	0.90 (0.09)	2.17 (0.17)
Montreal							
1 and 2	25 (5.7)	41 (9.3)	298 (67.4)	42 (9.5)	36 (8.1)	1.02 (0.17)	3.13 (0.47)
1 and 3	16 (6.3)	28 (11.0)	161 (63.1)	26 (10.2)	24 (9.4)	0.90 (0.15)	2.45 (0.36)
1 and 4	12 (8.0)	17 (11.3)	89 (59.3)	17 (11.3)	15 (10.0)	1.54 (0.66)	3.79 (1.57)
San Francisco							
1 and 2	79 (38.2)	31 (15.0)	43 (20.8)	17 (8.2)	37 (17.9)	3.01 (0.27)	3.80 (0.32)

[a] Level of discrepancy calculated by subtracting original measure from repeat measure excluding subjects who reported having had no sexual intercourse at the baseline interview.

[b] Mean difference and SE based on absolute values using all subjects in each cohort with follow-up interviews.

[c] Mean difference and SE based on absolute values using only subjects with discrepant responses between baseline and follow-up interviews.

[d] Third visit corresponding to the first interview following colposcopic examination.

Table 5 Summary statistics for the degree of repeatability in sexual history between baseline and first follow-up interviews for all cohorts combined, stratified by HPV and cervical lesion status

Variable	Stratification	Number of individuals (%) according to level of discrepancy between repeat and original measures[a]					Mean value (SE)[b]	Mean difference (SE)[c]	Intraclass Correlation Coefficient[d]
		≤-2	-1	0	1	≥2			
Age at first intercourse	None (all subjects)	589 (3.9)	2090 (13.7)	9953 (65.3)	2038 (13.4)	568 (3.7)	16.6 (0.07)	0.48 (0.02)	0.96 (0.001)
	HPV negative[e]	266 (4.7)	742 (13.1)	3647 (64.4)	749 (13.2)	251 (4.5)	17.5 (0.15)	0.55 (0.07)	0.95 (0.001)
	HPV positive	77 (4.7)	215 (13.1)	1049 (63.7)	237 (14.4)	68 (4.1)	17.1 (0.11)	0.56 (0.04)	0.91 (0.004)
	No lesions[e]	549 (3.9)	2002 (14.1)	9153 (64.5)	1940 (13.7)	539 (3.8)	17.0 (0.21)	0.49 (0.11)	0.96 (0.001)
	SIL/CIN or worse	31 (6.6)	56 (11.9)	295 (62.5)	69 (14.6)	21 (4.4)	17.4 (0.06)	0.62 (0.02)	0.89 (0.009)
Lifetime no. sexual partners[f]	None (all subjects)	376 (5.7)	580 (8.9)	4708 (72.0)	519 (7.9)	360 (5.5)	3.2 (0.14)	0.69 (0.09)	0.83 (0.004)
	HPV negative	268 (6.2)	387 (8.9)	3103 (71.3)	341 (7.8)	253 (5.8)	3.3 (0.30)	0.68 (0.15)	0.87 (0.004)
	HPV positive	75 (7.2)	114 (11.0)	657 (63.3)	116 (11.2)	76 (7.3)	3.7 (0.17)	0.79 (0.10)	0.90 (0.006)
	No lesions	334 (5.9)	524 (9.2)	4034 (71.1)	460 (8.1)	320 (5.7)	3.2 (0.36)	0.69 (0.79)	0.82 (0.004)
	SIL/CIN or worse	16 (5.3)	26 (8.7)	224 (74.7)	25 (8.3)	9 (3.0)	2.8 (0.15)	0.72 (0.10)	0.94 (0.006)

[a] Level of discrepancy calculated by subtracting original measure from repeat measure excluding subjects who reported having had no sexual intercourse at the baseline interview.

[b] Mean measure and SE reported at the baseline interview accounting for variance across cohort studies.

[c] Mean difference and SE based on absolute values using all subjects with follow-up interviews accounting for variance across cohort studies.

[d] Pooled intraclass correlation coefficient based on weighted z-transformed estimates with transformed SE using Taylor series approximation.

[e] Stratification by HPV status or cytology result based on baseline cervical specimen.

[f] Excluding data on no. of sexual partners from the Denmark cohort.

Table 6 Summary statistics for the degree of repeatability in sexual history between baseline and first follow-up interviews for all cohorts on the basis of categorical variables

Variable	Cohort designation	Identical coding (%)[a]	±One category discrepancy (%)[b]	±Two category discrepancy (%)[c]	Kappa (SE)
Age at first intercourse[d]					
	Denmark	7147 (84.3)	1322 (15.6)	14 (0.2)	0.76 (0.01)
	Costa Rica	2097 (83.7)	367 (14.7)	39 (1.6)	0.76 (0.01)
	São Paulo	1881 (86.0)	261 (11.9)	44 (2.0)	0.79 (0.01)
	Toronto	949 (92.4)	77 (7.5)	1 (0.1)	0.86 (0.02)
	Montreal	442 (92.9)	28 (5.9)	6 (1.3)	0.89 (0.02)
	San Francisco	175 (85.0)	31 (15.0)	0 (0.0)	0.71 (0.05)
Lifetime no. sexual partners[e]					
	Costa Rica	2113 (85.5)	338 (13.7)	16 (0.6)	0.77 (0.01)
	São Paulo	1959 (89.9)	214 (9.8)	8 (0.4)	0.83 (0.01)
	Toronto	816 (83.7)	156 (16.0)	3 (0.3)	0.76 (0.02)
	Montreal	404 (92.2)	32 (7.3)	1 (0.2)	0.89 (0.02)
	San Francisco	101 (63.5)	48 (30.2)	10 (6.3)	0.49 (0.05)

[a] Number of women providing equivalent responses for categorical variables at baseline and first follow-up interviews.
[b] Number of women providing responses differing by one category at baseline and first follow-up interviews.
[c] Number of women providing responses differing by two categories at baseline and first follow-up interviews.
[d] Categories based on typical coding excluding women reporting no sexual intercourse at first follow-up interview: ≤15, 16–17, 18–25, > 25 years.
[e] Categories based on typical coding excluding women reporting no sexual intercourse at first follow-up interview: 1, 2–4, 5–9, > 9 partners, (excluding data on no. sexual partners from the Denmark cohort).

Table 7 Logistic regression analysis of correlates of variation in reporting age at first intercourse and lifetime number of sexual partners

Variable	Category	Age at first intercourse				Lifetime number of sexual partners			
		Any variation[a]		Extreme variation[b]		Any variation[a]		Extreme variation[b]	
		OR[c]	95% CI	OR[c]	95% CI	OR[c]	95% CI	OR[c]	95% CI
Age	5 year intervals	1.36	(1.10–1.67)	1.42	(1.26–1.60)	1.09	(0.93–1.29)	1.07	(1.01–1.14)
Ethnicity	Caucasian	1.0	(ref)	1.0	(ref)	1.0	(ref)	1.0	(ref)
	African origin	1.32	(0.71–2.43)	1.78	(1.13–2.81)	0.99	(0.47–2.06)	1.35	(0.80–2.27)
	Asian origin	0.92	(0.61–1.41)	1.37	(0.54–3.50)	1.15	(0.76–1.75)	1.14	(0.52–2.52)
	Other	1.67	(0.91–3.08)	1.25	(0.81–1.92)	0.65	(0.31–1.37)	0.99	(0.58–1.67)
Marital status	Single	1.0	(ref)	1.0	(ref)	1.0	(ref)	1.0	(ref)
	Married	1.01	(0.73–1.39)	0.79	(0.58–1.05)	0.85	(0.61–1.19)	0.64	(0.47–0.89)
	Separated	0.84	(0.47–1.49)	0.55	(0.33–0.92)	1.24	(0.68–2.26)	0.94	(0.61–1.46)
Education	College/university	1.0	(ref)	1.0	(ref)	1.0	(ref)	1.0	(ref)
	Secondary	1.36	(0.75–2.48)	1.33	(0.84–2.10)	1.90	(0.84–4.27)	1.09	(0.65–1.82)
	Primary	1.21	(1.07–1.36)	2.14	(1.36–3.36)	1.94	(1.44–2.62)	1.77	(1.03–3.02)
Smoking status	Never	1.0	(ref)	1.0	(ref)	1.0	(ref)	1.0	(ref)
	Ex-smoker	0.83	(0.59–1.17)	1.16	(0.85–1.60)	1.27	(0.89–1.81)	1.60	(1.15–2.23)
	Current smoker	0.79	(0.56–1.10)	0.70	(0.50–1.00)	1.26	(0.89–1.77)	1.47	(1.03–2.10)
Interview interval	4 month intervals	1.03	(0.98–1.08)	0.95	(0.86–1.04)	1.02	(0.96–1.09)	1.04	(0.98–1.11)
Cohort	Costa Rica	1.0	(ref)	1.0	(ref)	1.0	(ref)	1.0	(ref)
	São Paulo	0.93	(0.78–1.10)	1.07	(0.69–1.68)	0.69	(0.53–0.89)	0.51	(0.31–0.83)
	Toronto	0.25	(0.11–0.58)	0.54	(0.26–1.12)	0.85	(0.42–1.72)	1.06	(0.57–1.98)
	Montreal	0.11	(0.04–0.27)	0.86	(0.34–2.16)	0.45	(0.22–0.93)	0.63	(0.29–1.39)
	San Francisco	0.30	(0.00–0.69)	2.47	(0.65–9.36)	0.85	(0.24–2.98)	2.57	(0.98–6.70)
	Denmark	0.95	(0.80–1.12)	0.31	(0.17–0.58)				
STD history	No	1.0	(ref)	1.0	(ref)	1.0	(ref)	1.0	(ref)
	Yes	1.08	(0.81–1.43)	0.85	(0.58–1.24)	1.09	(0.80–1.49)	1.39	(1.00–1.91)
Baseline HPV status	No	1.0	(ref)	1.0	(ref)	1.0	(ref)	1.0	(ref)
	Yes	0.84	(0.62–1.15)	1.10	(0.81–1.50)	1.59	(1.17–2.18)	0.65	(0.46–0.91)
Baseline cytology result	Normal/ASCUS	1.0	(ref)	1.0	(ref)	1.0	(ref)	1.0	(ref)
	SIL or worse	0.90	(0.38–2.13)	1.22	(0.75–1.97)	1.88	(0.78–4.55)	0.54	(0.26–1.12)
Time since first had intercourse	5 year intervals (at baseline interview)	0.99	(0.80–1.22)	0.79	(0.70–0.89)				

Table 7 Continued

Variable	Category	Age at first intercourse				Lifetime number of sexual partners			
		Any variation[a]		Extreme variation[b]		Any variation[a]		Extreme variation[b]	
		OR[c]	95% CI	OR[c]	95% CI	OR[c]	95% CI	OR[c]	95% CI
Lifetime no. sexual partners	(at baseline interview)					1.15	(1.11–1.19)	1.29	(1.25–1.33)
No. new sexual partners	(at first follow-up interview)					1.39	(1.25–1.55)	1.05	(1.01–1.09)

[a] Any discrepant response at the first follow-up interview or first interview following colposcopic examination (for Costa Rica).
[b] Discrepant response by ≥2 years at the first follow-up interview or first interview following colposcopic examination (for Costa Rica).
[c] Multivariate random effects model for binary responses mutually adjusted for all other variables shown.

partners, which may be due to the number of times the estimate for total number of partners had to be corrected for numbers of new partners to enable comparison with the baseline value. This distinction was also observed by Rohan *et al.*[19] using an earlier sample of women recruited in Toronto.

Under- or over-reporting of age at first intercourse could have occurred either at the baseline or at follow-up interviews, whereas there were further opportunities for error in reporting number of sexual partners, as study subjects were asked the total number of sexual partners reported at both interviews, as well as the number of new partners at the follow-up interviews. It is important to note that number of sexual partners has been a stronger predictor of HPV infection and cervical lesions than age at first intercourse. In the case of the SFSU cohort, women at baseline were likely to be very close in age to their first sexual experience and could accurately recall. However, by the time they were completing the second sentinel questionnaire they were much like the other cohorts in age and perhaps more likely to think in approximate terms. As the number of interim return visits increased, so did the number of opportunities for discrepant responses. With the accumulation of intermediate events and factors influencing reporting of sexual behaviour there can be substantial over- or under-estimation of relative risk estimates measured in long-term studies[29]. However, even after a relatively short interval of several weeks, the level of recall for Costa Rican women was similar to that for an interval of > 6 months. This would indicate that there was little influence of interval, though this may also be a result of the circumstances in which the interview and colposcopic exam were conducted.

In practice, the variability in responses may have less of an effect on relative risk estimates of HPV infection as outcomes, because sexual activity markers are typically analysed in categorical form in most epidemiologic studies of HPV and cervical cancer. The absolute component of variability that became irrelevant after collapsing the original variable codes into ordinal categories ranged from 9–25% (54–74% in relative terms) for age at first intercourse, and from 11–43% (42–76% in relative terms) for number of partners. The cut-points used were chosen arbitrarily for both variables, to represent typical categorisations used in previous studies from these cohorts[5,14,19,22–24]. Although we averaged out the degree of heterogeneity among cohorts to gain power, evaluation of effects across individual studies did not reveal contradictory conclusions. Odds ratios (OR) for baseline HPV ranged from 1.10 (95% CI 0.71–1.68) to 2.71 (95% CI 1.48–4.96) for any variation in reporting number of sexual partners, and from 0.17 (95% CI 0.03–0.84) to 1.06 (95% CI 0.51–2.18) for extreme variation.

Biased recall could be expected if women were made aware of their HPV infection or lesion status before their follow-up interview, given the potential emotional impact upon being informed of test results[30]. The pooled analyses indicated that age at first intercourse could be recalled with similar precision at the first follow-up interview, irrespective of HPV infection or lesion status. Interestingly, the latter variables (particularly HPV) were associated with variation in recall of number of sexual partners in an inconsistent fashion: positively for any variation and negatively for extreme variation. The association may not have resulted from recall bias due to knowledge of test result, since positive women were not explicitly targeted for follow-up, but rather from the expected differences between HPV-positive and HPV-negative women with respect to lifetime number of partners. It is conceivable that HPV-positive women may have had greater difficulty reporting frequencies (lifetime and new partners), simply because they engage more frequently in new sexual encounters than HPV-negative women.

In addition to ethnic and cultural factors and other characteristics that could not be accounted for in the analyses, the effect of cohort membership may indirectly have reflected the criteria used to select the follow-up sample in each study. For example, the Costa Rica and San Francisco cohorts targeted women with particular lesion outcomes, numerous sexual partners, and/or HPV positivity[23,26]. Cohort indicator was an important risk factor for reporting error. One key study-design characteristic that we hypothesised as a potential determinant of reporting accuracy was the method for collecting risk-factor information. Most studies relied on face-to-face, structured interviews via questionnaires, whereas the Montreal cohort used self-administered baseline and follow-up questionnaires that respondents completed on site. The latter study had the lowest reporting variation for age at first intercourse among all studies at 15%, a rate that varied little with time since the baseline interview. On the other hand, the Montreal cohort yielded error rates for lifetime number of sexual partners (32–41%, depending on follow-up return) that were comparable to those for the other Canadian university cohort (Toronto) and relatively high compared with São Paulo and Costa Rica.

Slightly different predictor profiles were observed for any and extreme reporting errors. In general, older age, lower educational attainment, and ethnicity of African descent were independent markers of lack of repeatability for sexual behaviour information. In addition, as would be expected, time since first occasion of sexual intercourse and multiplicity of partners were also strong, independent predictors of variability. It is worth remarking that these same characteristics serve as markers of

persistent HPV infection, a key precursor in the development of cervical precancerous lesions[26,31]. The increased propensity for misclassification of sexual behaviour information among these women may lead to bias in quantifying etiologic relations in the causal pathway leading to cervical lesion outcomes[3].

In conclusion, this collaborative study of repeatability of sexual history adds to the evidence that information on sexual behaviour markers is not reliably collected in epidemiological studies of sexually transmitted diseases. Even after controlling for study design differences and for potential confounding factors and events, cohort membership alone remains a substantial influence on the degree of reporting error. The impact of misclassification due to reporting error in sexual history on the epidemiological associations with HPV infection and cervical neoplasia remains to be investigated. The results from the present study can, however, be used in sensitivity analyses to illustrate the range of probable relative risk estimates that are consistent with scenarios in which reporting errors are accounted for. We are currently undertaking such an investigation.

Acknowledgements

The authors are indebted to Sven-Erik Kofoed (Danish Cancer Society, Copenhagen), Silvaneide Ferreira (Ludwig Institute for Cancer Research, São Paulo), Lisa Clayton (University of San Francisco), Vesna Stefanac (University of Toronto), and Jason Parente (McGill University, Montreal) for assistance with data management.

NF Schlecht is recipient of a pre-doctoral scholarship and EL Franco is recipient of a Distinguished Scientist Award, both from the Medical Research Council of Canada.

References

1 *Human papillomavirus*. International Agency for Research on Cancer Monographs on the Evaluation of Carcinogenic Risks to Humans: Vol. 64. Lyon: IARC, 1995.

2 Brinton LA, Hamman RF, Huggins GR *et al.* Sexual and reproductive risk factors for invasive squamous cell cervical cancer. *J Natl Cancer Inst* 1987;79:23–30.

3 Franco EL. The sexually transmitted disease model for cervical cancer: incoherent epidemiologic findings and the role of misclassification of human papillomavirus infection. *Epidemiology* 1991;2:98–106.

4 Schiffman MH, Bauer HM, Hoover RN *et al.* Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. *J Natl Cancer Inst* 1993;85:958–64.

5 Kjær SK, VanDenBrule AJ, Bock JE *et al.* Human papillomavirus — the most significant risk determinant of cervical intraepithelial neoplasia. *Int J Cancer* 1996;65:601–6.

6 Ley C, Bauer HM, Reingold A *et al.* Determinants of genital human papillomavirus infection in young women. *J Natl Cancer Inst* 1991;83:997–1003.

7 Rohan T, Mann V, McLaughlin J *et al.* PCR-detected genital papillomavirus infection: prevalence and association with risk factors for cervical cancer. *Int J Cancer* 1991;49:856–60.

8 Bauer HM, Hildesheim A, Schiffman MH *et al.* Determinants of genital human papillomavirus infection in low-risk women in Portland, Oregon. *Sex Transmit Dis* 1993;20:274–8.

9 Wheeler CM, Parmenter CA, Hunt WC *et al.* Determinants of genital human papillomavirus infection among cytologically normal women attending the University of New Mexico student health center. *Sex Transmit Dis* 1993;20:286–9.

10 Hildesheim A, Gravitt P, Schiffman MH *et al.* Determinants of genital human papillomavirus infection in low-income women in Washington, DC. *Sex Transmit Dis* 1993;20:279–85.

11 Kjær SK, DeVilliers EM, Caglayan H *et al.* Human papillomavirus, herpes-simplex virus and other potential risk factors for cervical cancer in a high-risk area (Greenland) and a low-risk area (Denmark): a second look. *Br J Cancer* 1993;67:830–7.

12 Franco EL, Villa LL, Ruiz A, Costa MC. Transmission of cervical human papillomavirus infection by sexual activity: Differences between low and high risk types. *J Infect Dis* 1995;172:756–63.

13 Kjær SK, VanDenBrule AJ, Bock JE *et al.* Determinants for genital human papillomavirus (HPV) infection in 1000 randomly chosen young Danish women with normal Pap smear: are there different risk profiles for oncogenic and nononcogenic HPV types? *Cancer Epidemiol Biomark Prev* 1997;6:799–805.

14 Richardson H, Franco E, Pintos J *et al.* Determinants of low- and high-risk cervical human papillomavirus infections in Montreal university students. *Sex Transmit Dis* 2000;27:79–86.

15 Franco EL. Statistical issues in studies of human papillomavirus infection and cervical cancer. In: Franco EL, Monsonego J, editors. *New developments in cervical cancer screening and prevention*. Oxford: Blackwell, 1997:39–50.

16 Saltzman SP, Stoddard AM, McCusker J *et al.* Reliability of self-reported sexual behavior risk factors for HIV infection in homosexual men. *Publ Health Rep* 1987;102:692–7.

17 Coates RA, Soskolne CL, Calzavara L *et al.* The reliability of sexual histories in AIDS-related research: evaluation of an interview-administered questionnaire. *Can J Publ Health* 1986;77:343–8.

18 Davoli M, Perucci CA, Sangalli M *et al.* Reliability of sexual behavior data among high school students in Rome. *Epidemiology* 1992;3:531–5.

19 Rohan T, McLaughlin J, Harnish DG. Repeatability of interview-derived information on sexual history: a study in women. *Epidemiology* 1994;5:360–3.

20 Kunin CM, Ames RE. Methods for determining the frequency of sexual intercourse and activities of daily living in young women. *Am J Epidemiol* 1981;113:55–61.

21 Coates RA, Calzavara L, Soskolne CL *et al.* Validity of sexual histories in a prospective study of male sexual contacts of men with AIDS or an AIDS-related condition. *Am J Epidemiol* 1988;128:719–28.

22 Moscicki AB, Palefsky J, Gonzales J, Schoolnik GK. Human papillomavirus infection in sexually active adolescent females: prevalence and risk factors. *Pediatr Res* 1990;28:507–13.

23 Herrero R, Schiffman MH, Bratti C *et al.* Design and methods of a population-based natural history study of cervical neoplasia in a rural province of Costa Rica: the Guanacaste project. *Pan Am J Publ Health* 1997;1:362–75.

24 Franco E, Villa L, Rohan T *et al.* Design and methods of the Ludwig-McGill Longitudinal study of the natural history of human papillomavirus infection and cervical neoplasia in Brazil. *Pan Am J Publ Health* 1999;6:223–33.

25 Munk C, Svare EI, Poll P *et al.* History of genital warts in 10,838 women 20 to 29 years of age from the general population. Risk factors and association with Papanicolaou smear history. *Sex Transmit Dis* 1997;24:567–72.

26 Moscicki A, Shiboski S, Broering J *et al.* The natural history of human papillomavirus infection as measured by repeated DNA testing in adolescent and young women. *J Pediatr* 1998;132:277–84.

27 Woolf B. On estimating the relationship between blood group and disease. *Ann Hum Genet* 1955;19:251–3.

28 Fleiss JL. The design and analysis of clinical experiments. New York: John Wiley and Sons, 1986:1–28.

29 Duffy SW, Rohan TE, McLaughlin JR. Design and analysis considerations in a cohort study involving repeated measurement of both exposure and outcome: the association between genital papillomavirus infection and risk of cervical intraepithelial neoplasia. *Stat Med* 1994;13:379–90.

30 Ramirez JE, Ramos DM, Clayton L *et al.* Genital human papillomavirus infections: knowledge, perception of risk, and actual risk in a nonclinic population of young women. *J Women's Health* 1997;6:113–21.

31 Hildesheim A, Schiffman MH, Gravitt PE *et al.* Persistence of type-specific human papillomavirus infection among cytologically normal women. *J Infect Dis* 1994;169:235–40.